Strategies for Mapping Concepts in Gastrointestinal Endoscopy Reports to the UMLS Metathesaurus

Michele Tringali[a], Thomas C. Rindflesch[b], Halil Kilicoglu[b], Marcelo Fiszman[b], Olivier Bodenreider[b]

[a]*Azienda Ospadeliera Santa Maria della Misericordia, Udine, Italy*
[b] *National Library of Medicine, Bethesda, MD, USA*

Abstract

Natural language processing methodologies are increasingly being used to support a wide range of clinical applications including medical decision support, outcome research, continuous quality improvement and information retrieval Our approach uses underspecified syntactic analysis and domain knowledge provided by the Unified Medical Language System® (UMLS).® We present three general strategies to improve mapping of strings in gastrointestinal endoscopy reports to Metathesaurus. We evaluate the effectiveness of our processing and then estimate the relative contribution of each strategy.

Keywords:

Natural Language Processing, Gastrointestinal System, Information Retrieval

Introduction

Automatic access to information in clinical free text can provide valuable support for medical informatics applications, including medical decision support, outcome research, continuous quality improvement, and information retrieval.

Natural language processing (NLP) provides a methodology for unlocking data from narrative reports, and a number of researchers have investigated various NLP techniques applied to clinical text. These systems explore the interaction of domain knowledge and linguistic structure. Due to the complexity of natural language, achieving highly accurate results requires NLP focused medical subdomain and type of report. The domain knowledge required is often constructed specifically for an application [1], [2] or is derived from marked-up training text [3].

We investigate techniques for minimizing the amount of effort required for a particular task, both with regard to NLP and the structured domain knowledge used. Recently, we used MetaMap [4] without modification or enhancement to map concepts in gastrointestinal endoscopy (GIE) reports to concepts un the UMLS Metathesaurus.® However, due to word sense ambiguity, synonymy, and context-specific meaning, useful results remain elusive [5].

In this paper, we discuss three strategies for improving on our preliminary results. As in the earlier work, we limit processing to the three semantic groups: Anatomy, Procedures, and Disorders. The first strategy, which we call domain processing, is implemented as an augmentation to MetaMap and addresses a small list of frequently occurring concepts important in GIE reports for which Metathesaurus coverage is not straightforward. This includes ambiguity in the domain, missing synonymy, and missing concepts. Such enhancement to MetaMap processing has been shown to be effective in other clinical domains [6].

We then call on two strategies, which focus on the available knowledge for the gastroenterology domain. The first of these is based on general methods of identifying domain-specific vocabulary in the Metathesaurus, and the second seeks to identify general medical terms that appear in GIE reports, but that are not specific to the gastroenterology domain. We use marked-up training text to discover such terms. We evaluate performance after these three strategies have been applied and estimated the relative contribution of each.

Background

The UMLS knowledge sources provide core support for this project. Semantic groups were devised to reduce the complexity of the Semantic Network [7]. Such groups organize 134 semantic types into 15 coarse grained aggregates. We constrained the mappings of GIE concepts to those belonging to the semantic groups Anatomy, Procedures, and Disorders.

Syntactic analysis to support the identification of semantic concepts in GIE reports begins with tokenization and lexical look-up in the SPECIALIST Lexicon [8]. Ambiguities are resolved by calling the Xerox Part-of-Speech Tagger [9], which is followed by an underspecified syntactic parse[10]. MetaMap processing then takes advantage of the underspecified parse to process only individual noun phrases.

MetaMap is a flexible program with many available options, one of which is "prefer multiple concepts" or not (default). This generates mappings that are either atomic or precoordinated. The reason to use one or the other depends on the application and circumstances. As an example, consider the term *esophageal varices* in the GIE domain. If multiple concepts are preferred in MetaMap, two concepts are produced: "Esophagus" and "Varicoce Veins." If multiple concepts are not preferred, the text maps to a single concept "Esophageal Varices."

Materials and Methods

Reports and Gold Standard

A sample of 80 esophago-gastroduodenoscopy reports from Clarkson University Hospital, Omaha, Nebraska, was randomly selected to be included in this study. Reports vary in length from 250 to 300 words. The sample was subsequently divided into training (55 reports, 15,653 words) and testing (25 reports, 6758 words) sets. A gold standard was generated in both samples by a board certified gastroenterologist (MT), who selected concepts along with their semantic types from the UMLS Metathesaurus (2003AA) to represent the gastrointestinal content of these reports. We then used the training set to devise general methods to improve mapping of free text to Metathesaurus. Baseline processing was performed with MetaMap set to prefer multiple concepts.

Domain Processing (D)

Processing the reports in the training set using MetaMap and evaluating them against the training gold standard shows some problems inherent in using the entire Metathesaurus for processing text in a particular domain. Errors were most frequently due to the domain-specific meaning of a particular concept. Our approach is to use the tagged training set iteratively to develop rules that resolved these problems in the GIE domain.

One example can be taken from the following sentence: "Examination of the **fundus** and cardia was likewise unremarkable". The string *fundus* maps to a general Metathesaurus concept "Fundi," which is defined as "The larger part of a hollow organ that is farthest away from the organ's opening. The bladder, gallbladder, stomach, uterus, eye, and cavity of the middle ear all have a fundus". It is clear that in this domain the meaning is "Gastric Fundus".

Another kind of context-related mapping occurs when a string is ambiguous between legitimate concepts, but one of them is more likely in this domain. For example, *dilation* maps to both "Endoscopic Dilatation" and "Pathological Dilatation." In a GIE report, "Endoscopic Dilatation" is more likely to be intended (although "Pathological Dilatation" is not impossible). Based on iteration in the training sample, we wrote rules that suppressed the less likely concept in such situations.

Missing synonyms in the Metathesaurus is another common problem that is seen, for example, in the sentence *the scope was advanced to the **mid esophagus***. "Middle third of esophagus" appears in the Metathesaurus but the synonym "mid esophagus" does not. Missing synonyms were added as they were discovered in the training set.

Errors due to missing concepts in the Metathesaurus were rare, but "Varices on varices" was noted in the training sample and added to a list specific to GIE reports.

Finally, we determined that higher accuracy could be achieved by setting MetaMap to "prefer multiple concepts" for all instances other than anatomical structures. Thus strings such as *esophageal erosions*, *gastrointestinal bleeding*, and *submucosal hemorrhage* were mapped to respective atomic concepts ("Submucosa" and "Hemorrhage," for example), while *distal esophagus, upper esophagus*, and *second part of the duodenum* were mapped to pre-coordinated concepts ("Descending portion of duodenum," for example).

Gastroenterology Domain Concepts (G)

Another method to improve accuracy is to limit the Metathesaurus concepts available to MetaMap to those in the gastroenterology domain. A list of gastroenterology concepts in the Metathesaurus was established for three areas of medicine corresponding to three semantic groups from the Semantic Network: Anatomy, Disorders, and Procedures. This was done by manually selecting one or more high-level concepts to be used as seed concepts. For example, we used the concepts "Esophageal Diseases" and "Gastrointestinal Diseases" for Disorders. We then extracted all descendants of the seed concepts by traversing recursively through the hierarchical relationships found in the Metathesaurus (namely, parent/child and broader than/narrower than).

In order to compensate for inaccuracies in the hierarchies, we further constrained the process by requiring that the descendants belong to the same semantic group as the seed concepts. We also eliminated some thesaurus relations not suitable to the domain. MetaMap processing was limited to Metathesaurus concepts in these lists.

Terms in Gastrointestinal Endoscopy Reports (T)

Limiting the concepts to gastroenterology with the above method is too strict. Some general medical concepts that are not specific to gastroenterology, such as "Veins," "Abdomen," "Erythema," "Findings," "Infection," and "Inflammation," are treated as false negatives by doing so. Such terms were identified after several iterations in the training set and were included as GIE terms.

Performance Metrics and Relative Contributions

We tested the effectiveness (precision and recall) of these three methods by comparing output to both the training and test sets. Effectiveness was measured for the three semantic groups combined (Anatomy, Procedures, and Disorders). An exact match was required for a concept to be considered correct.

We also measured the relative contribution of each of the methods proposed. In order to get an accurate estimate of the relative contribution of each method, we used the entire set of 80 reports as the basis for this determination.

Results

The effectiveness of our enhancements to MetaMap is showed in Table 1. The results are presented for the training and testing sets without enhancements to MetaMap (Base) and after the three methods were used to provide improvements (After).

Table 1 – Effectiveness measures for training and testing sets

	Base		After	
	Recall	**Precision**	**Recall**	**Precision**
Training	65%	64%	71%	79%
Testing	60%	64%	64%	80%

Discussion

We have presented three strategies to improve mapping free texts strings in gastrointestinal endoscopy reports to Metathesaurus. These strategies were added as enhancements to MetaMap. Results indicate that improved effectiveness persisted from the training set to the testing set (the two are disjoint).

The majority of the mistakes in the testing set were false-negatives and a consequence of the processing to distinguish atomic versus coordinated concepts. This processing is implemented by looking for the head of a noun phrase. If it is an anatomical concept, pre-coordination processing is pursued, otherwise atomic concepts are preferred. In the noun phrase *distal esophageal stenosis*, *stenosis* is the head and since it maps to a concept with (non-anatomical) semantic type 'Finding', the noun phrase is interpreted with atomic components "Distal," "Esophageal," and "Stenosis." However, the gold standard interpretation of this phrase is that *distal esophageal* are interpreted together as "Lower third of esophagus." ("Stenosis" is a separate concept in the gold standard.) Errors of these types also produce false positives, namely "Distal" and "Esophageal."

The relative contributions of each method based on the set of 80 reports are given in Figure 1. We use the convention that an upper case letter indicates the use of a particular method, while a lower case letter denotes its absence (D versus d, G versus g, T versus t). The baseline (dgt) has precision of .64 and recall of .63. Restricting to gastroenterology domain concepts alone (dGt) increases precision substantially to .78 but dramatically reduces recall to .21. Adding the GIE general

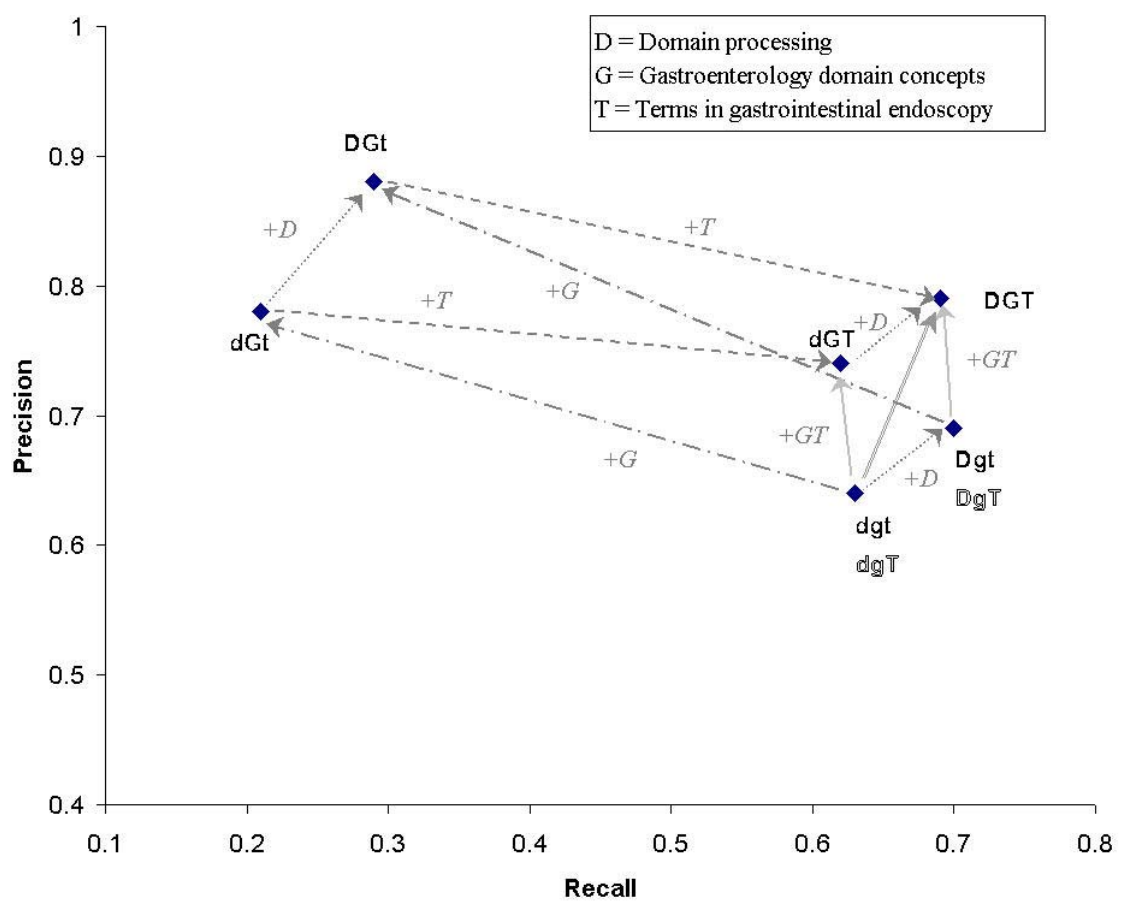

Figure 1 – Relative contributions

terms to the gastroenterology domain concepts (dGT) brings recall back to .62. Adding domain processing to these two strategies (DGT) produces the best results.

Notice that domain processing alone (Dgt) achieves a modest increase in precision (.69) and recall (.70) over the baseline. Simply restricting to gastroenterology concepts in addition to domain processing (DGt) improves precision considerably (.88) but reduces recall to .29.

One of the methods we used to improve results, using general medical terms in GIE reports (T), only shows improvement when coupled with the other two methods. That is, dgT and DgT are combinations that do not contribute an improvement. This is because such terms are in the Metathesaurus. It is only when they are eliminated through the use of the restricted gastroenterology concepts (G) that they must be reintroduced in order to improve recall.

One limitation of this project is the small number of reports used in the training set. The domain specific methodology depends on tagged reports in order to improve mapping performance. Another limitation is that only one GIE expert (MT), who is actively involved in the project, annotated the reports. It has been noted that inter-rater variability has an effect on evaluation in language processing [11].

There is a standard terminology for gastrointestinal endoscopy called MST® [12]. A version focused on structured reports has been integrated into the Metathesaurus [13]. A variant based on the information model of MST has also been used to create fully specified terms and is also included in the Metathesaurus (MTHMST). In our initial approach to selecting GIE domain concepts, we did not explicitly use this terminology, because we wanted to test methods that could be general enough for any domain.

In future work, we intend to use the information model provided by MST to support the identification of semantic predications in GIE free text reports. Long term goals of this project after semantic predications are extracted include automatic triggering of GIE related guidelines, semantic annotation of GIE images, information retrieval in the GIE domain and point of care decision support systems that depend on data from GIE reports.

Conclusion

We have developed and evaluated three general strategies for mapping free-text in gastrointestinal endoscopy reports into UMLS Metathesaurus concepts. These methods were used in conjunction with a mapping program called MetaMap. Preliminary results demonstrate improvements in the mapping process. The mapped concepts will serve as base for the interpretation of semantic propositions in GIE reports.

References

[1] Friedman C. A broad-coverage natural language processing system. Proc AMIA Symp 2000;:270-4.

[2] Hahn U, Romacker M, Schulz S. MEDSYNDIKATE--design considerations for an ontology-based medical text understanding system. Proc AMIA Symp 2000;:330-4.

[3] Christensen L; Haug PJ, Fiszman M. MPLUS: A probabilistic medical language understanding system. Proc of the Workshop on NLP in the Biomedical Domain, Association for Computational Linguistics 2002;:29-36.

[4] Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. Proc AMIA Symp 2001;:17-21.

[5] Tringali M, Fiszman M, Rindflesch TC, Bodenreider O. Automatic Extraction of Concepts from Gastrointestinal Endoscopy Reports. Proc of MIE, 2003.

[6] Rindflesch TC, Bean CA, Sneiderman CA. Argument identification for arterial branching predications asserted in cardiac catheterization reports. Proc AMIA Symp 2000;:704-8.

[7] McCray AT, Burgun A, Bodenreider O. Aggregating UMLS semantic types for reducing conceptual complexity. Medinfo 2001;10(Pt 1):216-20.

[8] McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. Proc Annu Symp Comput Appl Med Care 1994;:235-9.

[9] Cutting D, Kupiec J, Pedersen J, Sibun P. A practical part-of-speech tagger. Proceedings of the Third Conference on Applied Natural Language Processing 1992;:133-40.

[10] Rindflesch TC, Rajan J, Hunter L. Extracting molecular binding relationships from biomedical text. Proceedings of the 6th Applied Natural Language Processing Conference, Association for Computational Linguistics 2000 ;:188-95.

[11] Hripcsak G, Kuperman GJ, Friedman C, et al. A reliability study for evaluating information extraction from radiology reports. J Am Med Inform Assoc. 1999;6(2):143-50.

[12] Delvaux M, Korman LY, Armengol-Miro JR, Crespi M, Cass O, Hagenmuller F, Zwiebel FM. The minimal standard terminology for digestive endoscopy: introduction to structured reporting. Int J Med Inf 1998 Feb;48(1-3):217-25.

[13] Tringali M, Hole WT, Srinivasan S. Integration of a Standard Gastrointestinal Endoscopy Terminology in the UMLS Metathesaurus. Proc AMIA Symp 2002;:801-5.

Address for correspondence

Michele Tringali, M.D.

Azienda Ospedaliera Santa Maria della Misericordia

Udine, Italy